言语运动系统中前馈和反馈控制整合加工的作用机制

蔡笑 张清芳*

(中国人民大学心理学系，北京 100872)

摘 要 口语产生的最后阶段是发声运动，该阶段涉及到言语运动系统中前馈和反馈控制的整合加工。其中，前馈控制指个体自上而下地提取并执行产生目标语音的运动指令，而反馈控制指个体根据发声时产生的感觉反馈自下而上地调整言语运动，感觉目标和感觉预期是联系两者的重要枢纽。基于 DIVA(directions into velocities of articulators)神经计算模型，从语言习得和语言产生两个阶段，阐述了前馈和反馈控制整合的认知神经机制。在以往研究的基础上，重点梳理了听觉反馈如何帮助个体在线控制言语运动和更新前馈运动表征，以及 ERP 研究中 P1-N1-P2 成分波相应的认知内涵。此外，总结了影响讲话者前馈和反馈控制的各种因素，包括个体差异、训练经历和任务情境等，并提出这一领域应该重点关注的研究问题。

关键词 言语运动系统；前馈控制；反馈控制；听觉反馈

1 引言

言语是人际沟通最重要的方式之一，也是迄今人类掌握的最为精细复杂的运动技能之一。一般认为,言语产生可以分为三个过程：一是概念化(Conceptualization)，确立说话的意图和想表达的概念；二是言语组织(Formulation)，把要表达的概念转换为语言形式；三是发声(Articulation)，涉及到具体的语音和发音的计划(张清芳, 杨玉芳, 2003)。其中概念化和言语组织属于言语产生的计划过程，而发声属于言语产生的执行过程。当前心理语言学研究主要关注计划过程的认知机制(Levelt, Roelofs, & Meyer, 1999)，忽视了执行过程所涉及的复杂机制和理论阐述。两项言语产生的元分析表明，大量研究者将终端的发音执行等同于简单的、低层次的运动输出(Indefrey, 2011; Indefrey & Levelt, 2004)。事实上，将发音计划通过一系列发音器官的运动转化成可被感知和理解的声音序列是一个非常复杂的过程，涉及对发音运动进行预期、比较和控制等高级心理活动，这些心理活动的实现需要运动皮层、感觉皮层以及脑岛、小脑等(Golfinopoulos, Tourville, & Guenther, 2010)广泛的脑网络的参与。基于发音执行过程的复杂性，研究者提出也应该从运动控制角度进行研究，以弥补传统研究者从计划

收稿日期：2019-08-16

*北京市社科基金重点项目(16YYA006)，中国人民大学科学研究基金项目(中央高校基本科研业务费专项) (18XNLG28)项目，国家自然科学基金面上项目(31471074)资助，中国人民大学中央高校建设世界一流大学(学科)和特色发展引导专项资金支持。

通信作者：张清芳，E-mail: qingfang.zhang@ruc.edu.cn

过程切入的研究不足，为全面、系统地理解言语产生的心理机制提供理论和实践指导(Hickok, 2012)。

早期研究者比较关注一般性的运动控制过程，即个体如何产生精确的、以目标为导向的运动，强调感觉反馈(sensory feedback)在运动控制中的重要作用(Wolpert, Diedrichsen, & Flanagan, 2011)。近年来，研究者意识到口语产生本质上也属于精细化的运动控制，讲话者通过感觉反馈，尤其是更为重要的听觉反馈(auditory feedback)，对语音输出的准确性进行判断(Guenther, 2006; Guenther & Vladusich, 2012; Scheerer & Jones, 2012)。随着言语运动控制的探索不断深入，研究者逐渐认识到前馈控制(feedforward control)和反馈控制(feedback control)的认知内涵和作用机制。概括来讲，前馈控制指个体从言语运动系统中自上而下地提取并执行目标语音相应的运动指令，而反馈控制指个体自下而上地加工语言产生过程中实时产生的感觉反馈信息，并根据反馈信息中的言语错误调整和修正运动输出。两个系统并非独立起作用，而是相互合作，从而保证了言语产生执行过程的顺利进行(Guenther, Ghosh, & Tourville, 2006; Parrell, Lammert, Ciccarelli, & Quatieri, 2019; Perkell, 2012; Tourville & Guenther, 2011)。虽然越来越多的研究者开始从运动控制的视角关注言语产生的发生机制和神经基础(Cai, Ghosh, Guenther, & Perkell, 2010, 2011; Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Chen et al., 2015; Hickok, 2012; Scheerer & Jones, 2018)，但是相关研究还比较零散，缺乏对该领域的理论背景、研究热点和潜在问题的系统性论述。基于此，本文将从前馈和反馈控制整合加工的基础理论、研究现状以及影响因素等方面对言语运动控制进行讨论。

2 前馈和反馈控制整合的基础理论

言语运动控制不仅吸引着心理语言学和语音学等领域专家的兴趣，不少计算机和人工智能专家也致力于从神经计算的角度进行建模研究，其中最突出的是 Guenther 教授及其团队构建的 DIVA (directions into velocities of articulators)神经计算模型。该模型不仅关注个体在早期语言习得阶段如何形成前馈和反馈控制的能力，还重点阐述了成熟的讲话者在语言产生过程中如何整合前馈和反馈控制系统以确保正常的言语产生。

2.1 语言习得过程中前馈和反馈控制的整合

DIVA 模型提出，刚出生的婴儿不具备言语产生能力，语言习得需要经历咿呀学语和模仿两个发展阶段，期间均涉及了运动和感觉信息的整合过程。

咿呀学语阶段。在个体能够产生语音之前，必须首先建立起一般性声音(即非特异性语音的声音)的运动指令和对应感觉反馈之间的联结。在早期的咿呀学语阶段，婴儿开始尝试

发音，听觉和躯体感觉(体觉)皮层分别加工特定发音动作产生的听觉和体觉反馈信息，多次尝试后运动和感觉信息形成配对。此时，如果听觉或体觉错误发生时，个体可以根据配对后的运动-感觉联结将感觉错误信号转化为修正性的运动指令(Guenther et al., 2006; Tourville & Guenther, 2011)。

模仿阶段。当一般性声音的运动-感觉联结建立后，个体进入语音模仿阶段，开始学习特异性的语音产生。此时反馈系统提供的其他母语者产生的语音信息构成了语音模仿的模板，个体逐渐形成语音的听觉目标，即正确产生特定语音时期望出现的听觉反馈(Tourville & Guenther, 2011; Kearney & Guenther, 2019)。由于语音存在范畴化知觉现象，即某个语音在一定的变异范围内都被感知为该语音，研究者通常使用目标域来指代语音的听觉目标(Guenther & Vladusich, 2012)。Liu 和 Tian (2018)也指出听觉目标由听觉结果表征，并可以存储至听觉音韵系统。此外，通过反馈系统提供的自己正确发音时的体觉信息，包括发音器官(如唇和下颌等)的移动位置和不同发音器官的接触程度等(Parrell et al., 2019)，个体形成了体觉目标，即正确产生特定语音时期望出现的体觉反馈(Tourville & Guenther, 2011)。由此可见，反馈系统在语言习得初期至关重要，它可以帮助个体在记忆系统中形成母语语音的感觉目标表征。

在听觉目标习得后，个体开始基于模仿的机制学习产生语音的前馈运动指令(Guenther et al., 2006; Perkell, 2012; Tourville & Guenther, 2011)。起初，由于个体仍未在语音和相应运动指令之间形成稳定可靠的联结，发音运动的初始尝试产生的错误较大，言语运动完全依赖反馈控制系统。然而，随着每次发声运动，听觉控制系统对比自身实际产生的语音和语音的听觉目标，并依据听觉目标修正发音运动中的错误，最终准确习得语音的前馈运动指令。多次练习后，运动指令始终产生相同的感觉反馈并且未产生言语错误，强化了运动-感觉信息之间的联结，此时言语运动主要依赖前馈控制。综上，在早期语言习得阶段，前馈和反馈控制整合的核心是形成母语语音的感觉目标和产生语音的前馈运动指令以及建立运动指令和感觉反馈之间稳定的联结，逐渐发展出类似于成人言语运动控制技能。

2.2 语言产生过程中前馈和反馈控制的整合

语言产生过程中，成熟讲话者的言语运动控制更为复杂，涉及前馈系统和反馈系统以及两者联系的枢纽。DIVA 模型认为存储在记忆系统中的感觉目标为前馈控制和反馈控制的协同合作提供了接口，而其他的一些模型，例如任务动态(TD: task dynamics) (Saltzman & Munhall, 1989)、状态反馈控制(SFC: state feedback control) (Houde & Chang, 2015)等明确指出，通过内部前向模型(internal forward model)在线估算的感觉预期是联系前馈和反馈控制的

关键枢纽。事实上，虽然 DIVA 模型在理论框架中未引入内部前向模型，但它并未否认言语运动控制中存在内部前向机制(Guenther, 1995)。因此，研究者提出感觉反馈、感觉目标和基于运动指令的感觉预期共同促进了个体发展和保持精确言语产生的能力(Guenther, 1995; Hickok, 2012; Hickok et al., 2011; Houde & Nagarajan, 2011)。基于这三个因素之间的关系，我们将重点阐述前馈和反馈控制整合加工的作用机制。

开放环路的前馈控制。成熟讲话者的发声运动始于前馈控制系统。当个体计划产生语音集中特定音素或音节时，自上而下地提取该语音相应的前馈运动指令，运动指令发放至发音器官，指导上下唇、颌、舌以及喉部等详细的位置移动，最终实现发音(见图 1 中背景为灰色的部分) (Guenther, 2016)。前馈控制将运动指令输向发音器官，但不会把发声运动产生的感觉结果反馈回来影响当前的运动控制(Perkell, 2012)。这一过程类似于系统控制论中的开环控制(open-loop control)，是一种施控对象与受控对象只存在单向作用而无反馈信息的系统控制方式，因此，前馈控制实质上是从运动指令到声音输出的开放环路。在言语运动系统中，前馈控制存在明显的优势，因为个体可以迅速地发放运动指令，而且不需要处理复杂的感觉反馈，确保了流利地言语产生(Parrell et al., 2019)。但是，在母语或二语习得的早期，语音运动指令的准确性仍有待提高，开环控制系统中的运动指令一旦出错，必然会导致言语错误，因此前馈控制不适用于语言获得的早期阶段(Guenther, 2006; Guenther & Vladusich, 2012)。此外，当个体所处的环境不稳定时，容易发生一些干扰的情况，前馈控制不具备监测和修正语言输出中错误的能力，因此不适合不稳定的环境。

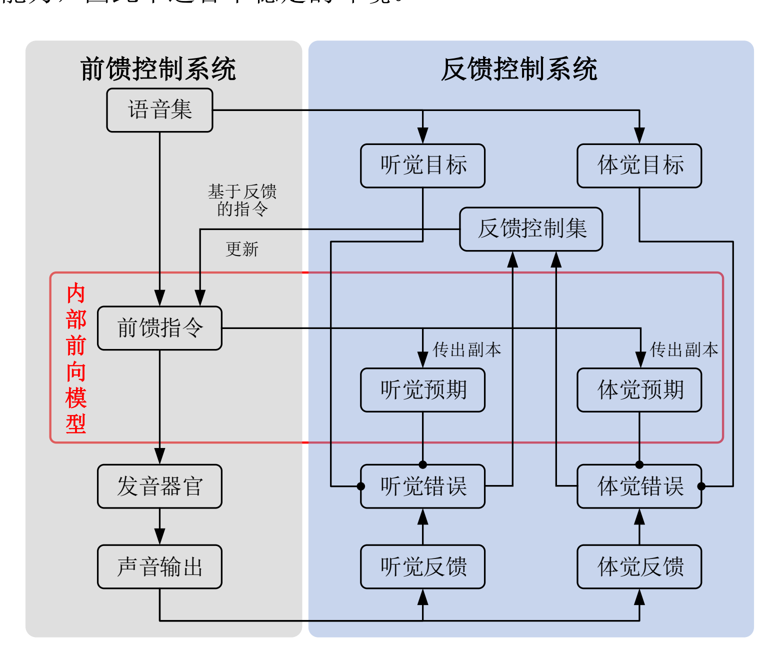


图 1 语言产生过程中的前馈和反馈控制系统

前馈和反馈控制联系的枢纽。前馈控制系统的输入信息是语音的运动指令，而反馈控制

系统的输入信息是发声运动产生的感觉反馈，两种性质的信息无法直接比较，因此言语运动控制模型需要明确前馈控制如何与反馈控制进行联系。研究者认为前馈控制系统同时存在感觉目标(记忆提取)和基于运动指令的感觉预期(motor-based sensory prediction) (在线估算)两种方式生成参与反馈控制的感觉表征(Liu & Tian, 2018; Tian & Poeppel, 2012)。

首先，DIVA 模型提出，每个语音激活运动指令的同时也会激活记忆系统中相应的感觉目标，而感觉目标编码正确产生语音时期望的感觉反馈，因此可以与实际的感觉反馈进行对比。在语言产生过程中，个体通过不断对比感觉目标和感觉反馈，判断发音动作的目标实现与否(Guenther, 2016)。实证研究和计算模型的证据也表明，感觉目标和实际感觉反馈之间的差异可以用于更新目标语音的运动指令(Guenther, 1994; Hickok, Houde, & Rong, 2011; Lametti, Krol, Shiller, & Ostry, 2014)。需要明确的是，虽然在图 1 的前馈控制系统中未包括感觉目标，但实际上从语音集到感觉目标自上而下的通路也属于前馈控制。

此外，高效的前馈控制可以通过内部前向模型提前预期感觉反馈(Franklin & Wolpert, 2011)，这种前瞻性使得前馈控制也被称为预期性控制(predictive control) (Parrell et al., 2019)。从认知层面来讲，感觉预期指个体在产生实际感觉反馈之前，通过心理内部估算当前声道运动状态以及随后可能的感觉反馈(Hickok, 2012; Tian & Poeppel, 2010, 2012)，这种预期完全依赖于语言习得阶段建立的语音运动指令和感觉输出之间的双向联结(Hickok, 2012; Tian & Poeppel, 2010, 2012, 2015; Tian, Zarate, Poeppel, 2016)。内部前向模型的神经基础是传出副本(efferent copy)，即在大脑内部复制的前馈控制系统向发音器官发放的运动指令，不同之处在于传出副本并不作用于发音器官，而是进一步传输至听觉皮层和体觉皮层，分别形成基于运动指令的听觉和体觉预期(见图 1 红色框) (Hickok, 2012; Hickok et al., 2011; Niziolek, Nagarajan, & Houde, 2013)。虽然图 1 的前馈控制系统中未包括感觉预期，但实际上从前馈指令到感觉预期自上而下的通路也属于前馈控制。预期性控制存在明显的速度优势，因为个体可以直接通过运动指令预期感觉反馈，即时衡量感觉目标的实现与否(Parrell et al., 2019)。但是，复杂的内部前向模型很难习得，当模型本身并不精确时，意味着预期的感觉反馈可能与实际的感觉反馈不匹配，此时如果没有反馈控制系统的参与，个体无法修正预期中的错误。

闭合环路的反馈控制。言语运动系统另外一个重要机制是反馈控制，其与前馈控制的本质区别在于它利用发声运动产生的感觉反馈进行言语运动控制(Parrell et al., 2019)。这一过程类似于系统控制论中的闭环控制(closed-loop control)，指一种将受控对象的结果输出反馈至施控对象，并对施控对象产生影响的系统控制方式，因此反馈控制本质上是一个从运动指令到错误信号，再从错误信号到运动指令的闭合环路。反馈控制涉及一系列加工过程(见图

1 中背景为蓝色的部分)：首先，感觉皮层编码由发声运动产生的感觉反馈；随后，个体将实际的感觉反馈与通过内部前向模型在线估算的感觉预期进行比较，在大多数情况下两者匹配，语言产生过程顺利结束，由于不需要修正当前的语音输出，此时感觉预期会抑制感觉目标的激活。但是在少数特殊情况下，感觉反馈和感觉预期不匹配，由于需要修正语音输出中的错误，此时感觉目标会充分激活编码感觉错误；最后，错误信号传输至反馈控制集，该模块基于感觉-运动之间的转换编码修正性的运动指令，当感觉错误多次发生时，基于反馈错误的修正性指令会更新当前语音的前馈运动表征(Guenther, 2016; Perkell, 2012)。由此可见，反馈控制的优势是适用于不稳定的环境，通过不断监测和修正言语输出确保了正常有效的语言交流(Hickok et al., 2011; Parrell et al., 2019)。但是它的速度劣势也十分明显，神经系统需要额外的时间加工感觉反馈信号并基于反馈做出合适的运动调整，因此言语运动控制如果过多地依赖感觉反馈，会导致运动缓慢和稳定性降低(Civier, Tasko, & Guenther, 2010; Guenther, 2006; Perkell, 2012)。

总之，前馈和反馈控制系统都具有各自的优势和不足，灵活的言语运动控制需要结合两种控制系统，这样既保留了前馈控制中的速度，也可以应对语言产生中的错误或外部意外的扰动(Guenther, 2016; Parrell et al., 2019)。基于上述加工过程可知，DIVA 模型中包括了很多的模块以及相应的参数，是一个庞大的流程模型。Guenther 教授也意识到当前的理论模型中包括很多自由参数，理应较好地拟合具体的实验数据，但是实证研究很难去验证模型中的假设。因此，Guenther 教授的团队正致力于构建简化的、可验证的三参数计算模型，模型仅包括听觉反馈控制增益(α_A: gains in auditory feedback control)、体觉反馈控制增益(α_S: gains in somatosensory feedback control)和前馈控制/学习速率(λ_{FF}: feedforward control/learning rate)三个参数，旨在评估语言产生过程中前馈和反馈控制系统的相对贡献(Kearney et al., submitted)。

2.3 言语运动控制的神经基础

参与语言产生的脑网络包括了双侧内外侧额叶皮层、顶叶皮层、颞上皮层、丘脑、基底神经节和小脑(Bohland & Guenther, 2006)，但仍然不清楚的是每个脑区在语言产生过程中具体的作用以及交互的方式。DIVA 模型基于大量 fMRI 研究，指出了参与前馈和反馈控制各个认知过程的脑区，为更全面了解言语运动控制提供了神经层面的结果。

DIVA 模型推测负责加工语音集的脑区位于左半球腹侧前运动皮层，主要包括腹侧中央前回、以及邻近的额下回后部和前脑区。前运动皮层参与读取常用语音精细编码后的发音运动程序，这些运动程序构成了言语运动控制中的前馈指令(Guenther, 2006, 2016; Guenther & Vladusich, 2012)。该假设的证据来自于大量言语运动障碍的脑损伤研究(Kearney & Guenther,

2019)，例如，研究者发现言语失用(apraxia of speech)病人在脑损伤发生前具备正常的语言产生能力，但是脑损伤(通常是腹侧前运动皮层区域)发生后，他们在运动计划和编码方面存在缺陷(Ballard, Tourville, & Robin, 2014; New et al., 2015)。

DIVA 模型强调了小脑的多种功能。第一，小脑对于学习和更新精细的前馈运动指令至关重要，其损伤很有可能会导致运动障碍(Ito, 2000)；第二，小脑的作用在于形成感觉目标(O'Reilly, Mesulam, & Nobre, 2008)；第三，小脑可能参与了反馈控制，因为小脑的活动与感觉错误的大小以及频率有关，而感觉错误的表征进一步驱动修正性的运动指令(Grafton, Schmitt, Van Horn, & Diedrichsen, 2008)。基于以上争议，Parrell 等(2017)对比小脑功能退化的病人和正常被试在前馈控制和反馈控制任务上的表现差异，发现小脑功能退化的病人保持精确前馈指令的能力下降，而反馈控制能力则与正常被试相似，支持小脑更关键的作用是更新和保持前馈运动指令。

神经影像研究发现，当实验者人为地改变返回到被试耳中的听觉反馈，会观察到颞叶区域显著激活，表明这一脑区与听觉错误的加工有关(Fu et al., 2006; Tourville, Reilly, & Guenther, 2008; Toyomura et al., 2007)。例如，Tourville 等(2008)的研究要求被试产生单音词(如"beck"和"bet")，命名过程中随机改变被试声学信号的第一共振峰(提高 30%或降低 30%)。实验者对比改变听觉反馈条件下和正常听觉反馈条件下大脑的激活情况，结果发现听觉错误显著激活颞上回后部和颞横回，其中激活峰值位于左侧颞横回后部末端。大量研究也证实颞横回负责监测听觉错误和编码听觉反馈和听觉预期不匹配的程度(Parkinson, Flagmeier, Manes, Larson, Rogers, & Robin, 2012; Zheng, Munhall, & Johnsrude, 2010)。Hickok (2012)在理论模型中提出颞平面负责听觉和运动信息的整合，因此该区域又被称为听觉-运动接口(Hickok et al., 2011; Hickok, Okada, & Serences, 2009)。

反馈控制主要激活右半球的脑区(Kalpouzos & Nyberg, 2010)，Tourville 和 Guenther (2011)首次在 DIVA 模型中引入了偏右侧化的反馈控制集，负责在言语错误发生时生成修正性的运动指令。早期的 DIVA 模型预测运动皮层的活动位于双侧，并主要集中于初级运动皮层(Guenther et al., 2006)。然而，Tourville 等(2008)发现人为改变共振峰时，听觉反馈异常会激活右侧前运动皮层，Golfinopoulos 等(2011)发现人为改变嘴唇或下颌移动位置时，体觉反馈异常同样会激活右侧前运动皮层。

综上，DIVA 模型从发展的角度阐述了个体在语言习得和语言产生阶段中前馈和反馈控制整合的加工机制。在语言习得阶段，个体建立语音的感觉目标表征和运动表征以及两者的联结。对于成熟的讲话者，前馈控制既可以直接提取目标语音的运动指令，也可以从记忆系

统中提取语音的感觉目标或通过内部前向模型形成运动指令相应的感觉预期，而反馈控制实时监测语言产生错误，并及时修正和更新前馈运动表征。从神经基础来看，言语运动系统涉及大脑双侧广泛的脑区，未来研究应致力于揭示更为精细的脑网络，以及不同脑区之间的联结模式。

3 言语运动控制的研究现状

早期研究者通常在行为层面探讨听觉反馈如何参与言语运动控制。近年来，随着电生理(如事件相关电位 ERP，一种在头皮记录的特殊脑诱发电位)和脑成像技术(如脑磁图 MEG 和功能核磁共振 fMRI)的发展，不少研究者结合行为与 ERP、MEG 和 fMRI 等技术探讨前馈和反馈控制的整合机制，并取得了丰硕的成果。

3.1 听觉反馈的功能：行为研究

当前研究者开始关注听觉反馈在语言习得和语言产生中的作用，提出听觉反馈具有三大主要功能：形成和保持语言运动技能、在线控制语言产生以及更新前馈运动表征(Cai, 2012; Guenther et al., 2006; Civier et al., 2010)。

听觉反馈在语言习得阶段的主要作用是形成语言运动技能。研究发现先天性耳聋的儿童无法获得流利的语言产生能力(Cowie et al., 1982)，患有轻度至重度听力损伤的学语前儿童通常无法产生可理解的语言(Oller & Eilers, 1988)，但是当借助人工耳蜗时，听力能力的恢复极大地促进了他们语言产生能力的习得(Tye-Murray & Spencer, 1995)。以上证据表明，具备正常的听力能力是习得语言运动技能的必备条件。即使是语言习得后，听觉反馈对于个体保持语言运动技能仍然具有重要作用。研究发现学语后听力丧失会导致个体言语运动的很多方面退化，包括语速、音强以及基频等(Lane & Webster, 1991)，但是人工耳蜗植入会促进语言可理解性的恢复(Gould et al., 2001)。

上述研究大多采用纵向研究方法，主要关注语言运动技能发生的缓慢变化，因而无法考察听觉反馈在实时言语运动控制中的作用。此外，母语者产生语音的运动指令已经高度自动化，通常听觉预期和听觉反馈匹配，因而很难考察基于听觉反馈错误的运动控制(Simmonds, Wise, & Leech, 2011)。早期研究者采用延时听觉反馈(delayed auditory feedback)范式，实验中利用特殊的设备和软件使讲话者的声音传到耳朵的时间有所延迟，此时人们通常会表现出类似口吃者的言语不流利现象。该范式利用延时方式干扰运动信息与听觉反馈的整合，证明语言产生会受到听觉反馈的影响(Mitsuya, Munhall, & Purcell, 2017; Tian & Poeppel, 2015)。但是有研究者质疑延时明显地改变了听觉反馈的自然模式，被试能够意识到改变，因此提倡

更为精细的实验操纵(Cai, 2012)。近年来，现代声学信号处理技术革新了言语运动控制的研究手段，催生出听觉反馈扰动(auditory feedback perturbation)范式(Cai, Beal, Ghosh, Tiede, Guenther, & Perkell, 2012; Cai et al. , 2010)。由于听觉反馈表现为外显的声学信号，容易获得和操纵，研究者开始以精细控制的方式实时改变听觉反馈中的多种声学参数，其中最常用的包括音强(Bauer, Mittal, Larson, & Hain, 2006; Patel, Reilly, Archibald, Cai, & Guenther, 2015)，基频(Chang et al., 2013; Franken et al., 2018a, 2018b)和共振峰(Cai et al., 2012; Daliri, Wieland, Cai, Guenther, & Chang, 2017)。扰动范式最主要的优势是建立了听觉反馈感知和言语运动调整之间的因果关系。研究者主要关注的问题包括：

第一，当听觉反馈扰动发生时，被试如何调整言语运动？目前大量研究发现，当听觉反馈扰动的幅度未超出阈限(即仍可能认为声音由自己发出)，讲话者通常会向人为扰动的相反方向修正言语运动，这种现象被称为补偿性反应(compensatory response)。例如，当反馈到耳朵里的声音基频被提高，讲话者会自动降低基频(Chang et al., 2013)。这种相反方向的补偿性反应表明言语运动控制的目的是使听觉反馈维持在稳定状态。

第二，被试基于听觉反馈错误的运动调整发生的时间进程如何？以及前馈和反馈控制的整合过程是自动化加工或是控制性加工？研究发现当听觉反馈扰动发生后，被试调整言语运动的过程十分迅速，通常在扰动发生后的 100~200 ms 内(Bauer et al., 2006; Franken et al., 2018a; Cai et al., 2012)。Scheerer 和 Jones (2018)发现反馈控制系统即使对语言产生中微小的偏差也敏感，实验后的调查也表明被试有时意识不到扰动和补偿性反应的存在(Cai et al., 2011; Parrell et al., 2017)。Munhall 等(2009)在实验中扰动共振峰并告知被试扰动的具体操纵，要求被试不要进行补偿性反应。结果发现即使告知被试后，他们仍然会向扰动的相反方向调整运动，表明补偿性反应不易受有意识的策略影响。以上证据(即发生时间快、无意识的修正以及不受策略影响)都支持前馈和反馈控制的整合过程是自动化的加工(Korzyukov et al., 2012a)。

第三，听觉反馈扰动的幅度与言语运动调整的幅度之间的关系及其认知机制是什么？研究者发现补偿性反应的幅度通常只占听觉反馈扰动幅度的小部分，例如 100 音分的音高听觉反馈扰动只能诱发小于 50 音分的补偿性反应(Korzyukov et al., 2012a; Scheerer et al., 2013b), 甚至有研究报告只有 8 音分左右的补偿性反应(Scheerer et al., 2013a)。一方面，这是由于言语运动控制同时依赖前馈控制和反馈控制系统(Tourville & Guenther, 2011)。另一方面，由于听觉反馈扰动范式只是造成了预期的听觉反馈和实际的听觉反馈的不匹配，而预期

的体觉反馈和实际的体觉反馈匹配，因此反馈控制系统需要同时加工紊乱的听觉反馈信息和吻合的体觉反馈信息。这种扰动和补偿的不对称现象说明，听觉反馈不是言语运动系统中唯一的信息输入(Lametti et al., 2012)，前馈控制和体觉反馈也参与言语运动控制。

听觉反馈的第三个作用是更新前馈运动表征，Delvaux 和 Soquet (2007)发现成人语言产生仍会受到周围语言环境的影响，人们会根据语言环境调整语音模式(例如音高、元音特征)。这说明前馈指令建立后不是固定不变的，而是会受到听觉反馈的影响不断校准更新。研究者发展出言语适应(speech adaptation)范式，考察长时程的听觉反馈扰动如何影响前馈运动表征的更新(Cai et al., 2010; Daliri et al., 2017; Parrell et al., 2017)。目前该范式被广泛应用于基频和共振峰控制的研究，典型的言语适应范式包括四个连续的阶段(Daliri et al., 2017)。1)基线期(baseline phase)：人们在正常听觉反馈条件下发声；2)渐变期(ramp phase)：在听觉反馈扰动条件下发声，听觉反馈扰动的幅度逐渐变化直至到达峰值；3)保持期(hold phase)：持续在最大幅值的听觉反馈扰动条件下发声，通过多个试次重塑语音的运动表征；4)结束期(end phase)：最后在正常听觉反馈条件下发声。研究发现保持期与基线期相比，讲话者会向扰动的相反方向做出补偿性反应，也称适应性反应(adaptive response)，当结束期扰动撤销后适应性反应仍然会保持短暂时间，也称遗留效应(aftereffect) (Daliri et al., 2017; Parrell et al., 2017)。基线期和结束期被试都接受正常的听觉反馈，结束期的遗留效应说明被试已经更新特定语音的前馈运动表征。这表明人为造成的听觉错误会引发反馈控制系统修正运动，从长时程来看，修正性指令被纳入前馈控制系统指导随后的发音行为。

研究者也关注言语适应能否泛化到其他未扰动的语音上。Cai 等(2010)考察汉语母语者在面对共振峰扰动时如何产生三元音/iau/，以及适应性反应能否泛化到其他时间或空间特征与/iau/不一样的元音上，如/uai/, /ia/和/au/等。结果发现泛化模式广泛存在，但是比较微弱，且随着两个语音相似程度的降低而递减。Reilly 和 Pettibone (2017)也发现重复扰动特定元音的听觉反馈不仅会改变被扰动元音的产生(适应性反应)，也会改变附近未被扰动元音的产生(泛化)。以上证据表明不同元音的前馈运动指令不是独立表征的，否则某个元音基于听觉反馈错误的前馈表征更新不会影响其他元音的产生。因此，研究者推测时间和空间特征相似的元音可能共享某些机制以负责计算语音的运动轨迹，这样一个元音运动-感觉映射的调整会导致其他元音运动编码的改变。这对 DIVA 模型提出了挑战，DIVA 模型将不同语音视为独立的个体，单独存储前馈运动指令(Guenther et al., 2006; Tourville & Guenther, 2011)，因此不能解释不同元音中发现的泛化反应。进一步的研究需要解释不同语音间的泛化，以完善

DIVA 模型。

3.2 前馈和反馈控制整合加工的时间进程：ERP 和 MEG 研究

虽然大量研究从行为层面揭示言语运动控制的发生机制，但是行为研究只能依托外显的运动调整，无法考察听觉反馈错误监测的过程。随着认知神经科学的兴起，不少研究者采用听觉反馈扰动范式，并结合 ERP、MEG 等高时间分辨率技术探讨前馈和反馈控制整合加工的时间进程(Behroozmand et al., 2016; Chen et al., 2012a, 2012b; Franken et al., 2018b; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Scheerer et al., 2013a, 2013b; Scheerer & Jones, 2014, 2018)。研究者关注听觉反馈扰动幅度与神经活动变化的关系，发现了反映言语运动控制的 N1-P1-P2 成分波(Chen et al., 2012b; Liu et al., 2011; Scheerer et al., 2013a)。

首先，研究者普遍认为 P1 成分(MEG 实验中对应 M50 成分)反映了被试对早期听觉刺激变化的监测。例如，Scheerer 等(2013a)在记录 ERP 的同时随机扰动听觉反馈，结果发现 P1 指标只对是否扰动敏感，但是对扰动的幅度不敏感。Korzyukov 等(2012a)也发现 P1 的诱发是以全或无的方式，反映了对一般性听觉刺激的监测。

相比于 P1 成分, 早期研究者更加关注语言特异性的 N1 成分波(MEG 实验中对应 M1 或 M100 成分)以及言语运动诱发的抑制现象(SIS: Speech-induced suppression)。SIS 具体表现为：比较被试发声时听自己正常的听觉反馈与自己受到扰动的听觉反馈(Heinks-Maldonado et al., 2005, 2006)，或比较被试发声时听自己正常的听觉反馈与听别人的声音(Heinks-Maldonado et al., 2005, 2006)，或比较被试发声时听自己正常的听觉反馈与未发声时被动地听自己声音的录音(Houde, Nagarajan, Sekihara, & Merzenich, 2002)，N1/M1 波幅均会降低。SIS 现象也得到 fMRI 证据的支持，研究发现被试发声时听自己正常的听觉反馈与听自己受到扰动的听觉反馈相比，听觉皮层的活动降低(Parkinson et al., 2012; Zheng et al., 2010)。因此，SIS 本质上反映了通过内部前向模型预期感觉反馈的机制(Behroozmand et al., 2016; Heinks-Maldonado et al., 2005; Kort, Nagarajan, & Houde, 2014)，当预期的听觉反馈与实际的听觉反馈匹配时，诱发出 SIS 现象；当两者不匹配时，听觉皮层需要额外的资源加工言语错误。SIS 也被认为是讲话者区分自己和外部产生言语的重要机制。

研究者认为 N1 波幅的变化反映了个体对听觉反馈扰动的监测以及对内部和外部产生言语的区分(Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Liu et al., 2011; Scheerer et al., 2013a)。在 Scheerer 等(2013a)的研究中，任何听觉反馈扰动条件都比未扰动条件诱发更大的 N1 波幅，并且波幅变化受到扰动幅度的影响，具体表现在大的扰动(400 音分)比相对较小的扰动(50-250 音分)诱发更大的 N1 波幅。这与 Liu 等人(2011a)的研究结果吻

合，500 音分和 200 音分扰动比 100 音分扰动诱发更大的 N1 波幅(同见 Behroozmand & Larson, 2011)。因此，研究者推测 N1 成分实际反映了讲话者对比听觉反馈和听觉预期的过程。当实际的听觉反馈违反了预期，但仍可以被认为是由自己产生的言语时，N1 波幅的增加是以全或无的方式；但当大的听觉反馈扰动发生时，讲话者认为超出了生理学上自身可产生声音的范围，因而判断听觉反馈来自他人，并诱发更大的 N1 波幅。N1 的潜伏期可能反映了听觉反馈加工的效率，有研究表明随着扰动幅度的提高，N1 潜伏期缩短，说明被试更快速地监测到大的听觉反馈错误(Liu et al., 2011; Scheerer et al., 2013a)。

此外，研究者开始越来越关注基于反馈错误的运动控制在神经层面上的表现，并发现了 P2 成分波(MEG 实验中对应 M2 或 M200 成分)。与 N1 相比，P2 波幅根据扰动幅度的大小表现出最系统的变化。Scheerer 等(2013a)发现在 0 到 250 音分扰动区间内，P2 波幅随扰动幅度的增加而增加，但 P2 波幅并非一直线性上升，当扰动幅度大于 250 音分时，波幅开始下降。值得注意的是，行为层面上的补偿性反应也表现出与 P2 波幅变化相似的模式。当听觉反馈扰动较小时，补偿性反应随扰动幅度的增加而增加；但是当大的听觉反馈扰动发生时，讲话者判断听觉反馈来自他人而非自己，因而不会做出补偿性反应或补偿性反应的幅度较小(Scheerer et al., 2013a; Tian & Poeppel, 2015)。此外，行为-脑电的相关分析和回归分析都表明行为上的补偿性反应和 P2 波幅存在正相关，因此研究者认为 P2 可能反映了听觉错误的计算以及相应的运动指令修正(Chen et al., 2015; Jones, Scheerer, & Tumber, 2013; Kort et al., 2014; Scheerer et al., 2013a; Scheerer & Jones, 2014)。

少数研究者开始尝试利用时频分析探讨大脑如何整合运动和听觉反馈信息进行言语运动控制，发现 θ 频段和 δ 频段的震荡活动具有重要意义(Behroozmand et al., 2015; Cavanagh & Frank, 2014; Cruikshank, Singhal, Hueppelsheuser, & Caplan, 2012)。例如，Cavanagh 和 Frank (2014)发现 δ 频段(1 到 4 Hz)和 θ 频段(5 到 8 Hz)的活动反映了对扰动听觉反馈中新异、冲突以及错误信息的加工，因此在神经层面标记了对认知控制的需求。Behroozmand 等(2015)进一步对比音乐家和非音乐家在音高听觉反馈扰动实验中神经活动的差异。研究发现，被试对听觉反馈扰动做出补偿性反应的同时，伴随相位一致性 θ 频段额中区域的震荡活动，且音乐家 θ 频段的震荡强度高于非音乐家。在扰动呈现 1 秒后，额叶区域出现非相位一致性 δ 频段的震荡活动，且音乐家 δ 频段的震荡强度低于非音乐家。此外，δ 频段的震荡活动与扰动结束后被试重新调整音高以回归基线水平的能力有关。研究者据此推测 θ 频段在神经生理层面标记音乐家提高的音高加工能力，反映了人类整合听觉反馈信息以控制言语产生的机制(Behroozmand et al., 2015; Cruikshank et al., 2012)。而 δ 频段标记言语适应机制，即个体如何

依据听觉反馈错误更新前馈运动指令并指导后续言语产生(Behroozmand et al., 2015)。时频研究从不同角度补充了前馈和反馈控制整合加工的神经机制，未来研究应继续探讨神经震荡活动的认知涵义。

综上，P1，N1 和 P2 波形的变化表明，当听觉反馈扰动发生时，个体可以在早期监测到听觉信息的异常，调用更多的认知资源加工听觉错误，并基于反馈信息进行言语运动的调整，θ 频段和 δ 频段可能参与了运动-感觉信息的整合过程。

4 影响前馈和反馈控制整合加工的因素

言语运动系统的正常运转依赖于前馈和反馈控制的协同合作(Guenther et al., 2006; Perkell, 2012; Tourville & Guenther, 2011; Hickok, 2012)，因此两者在言语运动控制中的相对贡献及其影响因素是非常重要的研究问题。研究者通过观察个体受听觉反馈扰动干扰的程度来考察前馈和反馈控制的相对权重，实验逻辑是如果被试对前馈控制的依赖程度越低，则会对感觉反馈的依赖程度越高，相应地更容易受到听觉反馈扰动的干扰。大量研究以补偿性反应或 P1-N1-P2 成分波为切入点，发现影响因素主要表现在三方面：第一，个体差异，包括年龄、性别、发声变异性和言语障碍等；第二，训练经历，包括语言经历和音乐经历等；第三，任务情境，包括可预期性和注意负荷等。

4.1 个体差异

前馈和反馈控制的表现存在个体差异，研究者关注产生个体差异的本质原因，尤其是语言产生障碍群体的运动控制缺陷。

年龄和性别。语音表征和运动表征的整合开始于婴儿的咿呀学语，并在随后的语言习得过程中一直保持可塑性以适应发音器官的生长、肌肉组织的增加以及肺活量的变化(Guenther et al., 2006)。有研究关注儿童和成人听觉反馈控制的差异，例如 Liu 等(2010a) 发现 7 到 12 岁讲英语的儿童做出补偿性反应的潜伏期长于成人。Scheerer，Liu 和 Jones (2013b) 的横断研究考察了 4 到 30 岁年龄跨度的被试，结果发现补偿性反应和 P1-N1-P2 波幅均受到年龄的影响。对正在习得语言的儿童来说，听觉反馈可以帮助他们建立前馈表征，相应的权重会被提高，但当发展停止时，前馈表征保持相对稳定，听觉反馈提供的信息则变得冗余，因此提高前馈控制的权重可以增加言语流畅度和减少外界的干扰(Civier et al., 2010)。有证据表明，听觉反馈扰动实验中女性比男性产生更小的补偿性反应，且 N1 和 P2 潜伏期更短(Chen et al., 2010; Swink & Stuart, 2012)。Li 等(2018)发现青年男性比青年女性诱发更大的 N1 和 P2 波幅。研究者认为可能是两性间的生理差异导致了前馈和反馈控制的差异(Chen, Liu, Jones,

Huang, & Liu, 2010; Kakimoto et al., 2016)

发声变异性。在音高扰动任务中，实验者通常要求被试进行持续元音发声任务，发声变异性指个体在基线条件下，即未扰动听觉反馈时音高变化的标准差。部分研究者关注发声变异性的个体差异如何影响听觉反馈控制(Scheerer & Jones, 2012; Scheerer et al., 2013a)。Scheerer 和 Jones (2012)发现发声变异性更大的被试对听觉反馈扰动产生更大的补偿性反应，揭示言语输出的不稳定会导致个体增加对听觉反馈输入信息的依赖性，以维持稳定的言语运动控制。

言语障碍。基于已有神经影像研究，DIVA 模型在大脑中标记出不同言语障碍的脑损伤区域(Guenther, 2016)。言语障碍的严重性取决于脑损伤区域是影响了前馈控制系统或是反馈控制系统。在语言习得的发展过程中，反馈控制系统对于形成前馈指令是不可或缺的。但是，当语言习得后，前馈控制系统可以在几乎不需要反馈控制系统参与的情况下发放运动指令并产生语音。因此，对于成熟的讲话者而言，负责反馈控制系统的脑区受损对语音输出的影响相对有限。但是，负责前馈控制系统的脑区受损将引起明显的言语运动障碍(Kearney & Guenther, 2019)，常见的语言产生障碍如言语失用、口吃(stuttering)、构音障碍(Dysarthria)等得到研究者的广泛关注。

言语失用是一种运动计划编码障碍，其典型特征是语速慢、语音扭曲、韵律异常等。言语失用患者的脑损伤主要位于左半球额下区域，尤其是腹侧前运动皮层，该区域负责提取精细编码的前馈运动指令(Guenther, 2006, 2016)。因此，言语失用主要是由于前馈运动控制系统的缺陷(Kearney & Guenther, 2019; Tourville & Guenther, 2011)。研究者提出额下区域的损伤可能也会影响个体提取语音的感觉目标，而反馈控制的运作机制是对比感觉目标和实际的感觉反馈，因此该脑区损伤可能会进而导致反馈控制受损(Ballard et al., 2018; Kearney & Guenther, 2019)。但是，反馈控制缺陷仍未得到实证研究充分的验证。Maas 等(2015)利用噪音掩蔽(noise masking)范式考察了言语失用的本质，结果表明病人前馈控制受到破坏，导致反馈控制扮演更突出的角色(同见 Iuzzini-Seigel, Hogan, Guarino, & Green, 2015)，该结论也得到计算机模拟证据的支持(Terband, Rodd, & Maas, 2015)。

口吃是一种语言流畅性障碍，其典型特征是语音重复、语音延长和沉默。DIVA 模型指出，口吃者比正常被试更高地激活参与反馈控制集的主要脑区(右半球的中央前回和额下回区域)，而反馈控制集负责计算基于反馈错误的修正性运动指令，因此口吃主要是由反馈控制的缺陷或异常(Tourville & Guenther, 2011)引起的。在行为层面，Cai 等(2012)通过随机扰动听觉反馈也证实口吃者听觉-运动转换的功能存在缺陷(Daliri et al., 2017)。另外一种观点

认为，口吃者从前馈控制系统中直接提取运动指令的能力受损，导致他们过多地依赖听觉反馈进行运动控制，影响了语言产生的流畅性(Civier et al., 2010)。在电生理层面，Daliri 和 Max (2015a, 2015b)证实了口吃者预期听觉反馈的一般性能力存在缺陷，因此其前馈控制系统可能受损。

4.2 训练经历

语言经历。研究发现声调语言的经历会减少个体受音高扰动影响的程度，这是由于声调语言可使用音高变化来区别意义，而非声调语言通常只用音高变化改变语气，因此不同语言的使用者会发展出语言特异性的音高控制能力(Chen et al., 2012b; Liu et al., 2010b; Ning, Shih, & Loucks, 2014; Ning, Loucks, & Shih, 2015)。英语是典型非声调语言，而汉语是典型的声调语言，Ning 等(2014)首次对比有无声调语言经历对音高控制的影响，并将研究问题扩展到第二语言学习领域(汉语二语学习者)。结果发现，汉语母语者产生幅度最小的补偿性反应，而二语学习者的表现模式介于汉语母语者和英语母语者之间，揭示语言训练经历对音高前馈和反馈控制的塑造。Liu 等(2010b)对比粤语和汉语母语者听觉反馈控制的差异，虽然两种语言都属于声调语言，但粤语的声调系统(6 声)比汉语(4 声)更为复杂精细，结果发现粤语母语者比汉语母语者产生更小的补偿性反应。以上研究都证实精细的声调系统使讲话者拥有稳定的音高前馈控制能力，因而不易受听觉反馈扰动的干扰。

音乐经历。研究者发现音乐训练影响了音高的前馈和反馈控制能力。当主试在持续元音发声任务中随机扰动听觉反馈的音高，有经验的音乐家比非音乐家产生更小的补偿性反应，说明音乐训练提高了音高控制的稳定性，减少了来自外部反馈的干扰(Jones & Keough, 2008; Keough, Hawco, & Jones, 2013)。Ning 等(2015)在听觉反馈扰动任务中对比汉语母语者、汉语二语学习者、经过训练的歌手和未接触过声调语言的讲话者四组被试的表现，发现汉语母语者在元音发音和汉语声调任务中均表现出最稳定的音高控制，经过训练的歌手也比未接触过声调语言的被试在元音发声任务中表现出更稳定的音高控制，但在语言特异性任务中未发现差异。以上结果说明控制声调/音高的经历有助于形成更稳定的前馈运动表征，但由音乐训练提高的音高前馈控制能否泛化到语言任务中还有待验证。

4.3 任务情境

可预期性。特定事件稳定出现的频率越高，预期事件出现的能力就越强。在言语运动控制中，前馈控制可以通过内部前向模型预期听觉反馈，因此研究者猜测如果以可预期的方式扰动听觉反馈，被试可能会提高前馈控制的权重，进而减少来自反馈的干扰。该假设得到大量证据的支持(Chen et al., 2012a; Korzyukov et al., 2012b; Scheerer & Jones, 2014)。Korzyukov

等(2012b)发现扰动方向的可预期性诱发更小的 N1 波幅；Scheerer 和 Jones (2014)发现扰动幅度的可预期性导致补偿性反应的幅度变小，N1 波幅减小，补偿性反应和 N1 潜伏期加快。这些发现证实可预期性提高了前馈控制系统的权重。

注意负荷。日常生活中, 讲话者接收听觉反馈的同时可能还需要加工其他模态的信息，但注意资源是有限的，因此有研究者提出注意负荷可能影响听觉反馈加工。Tumber, Scheerer 和 Jones (2014)分别在单任务和双任务情境下随机扰动听觉反馈，单任务情境要求被试被动地观看视觉线索发声，双任务情境要求被试在发声的同时辨别字母串以增加注意负担。结果发现，单任务比双任务情境产生更大的补偿性反应，说明当被试的注意被分散时，更少的注意用于加工听觉错误。Liu 等(2015)进一步验证了该观点，他们发现当被试选择性注意听觉反馈时，会产生更大的补偿性反应。然而，Alsius, Mitsuya 和 Munhall (2013)的研究未发现补偿性反应在注意集中和注意分散任务上的差异。对于矛盾的结果，未来需要更多的研究来探索注意调控听觉反馈控制的机制。

5 总结与展望

理解前馈和反馈控制系统如何在大脑中协同合作以保证正常的言语产生具有重要的理论意义和实践价值。近十年，研究者致力于构建细致的言语运动控制基础理论和神经网络，尤其关注个体如何利用听觉反馈在线调整言语运动以及更新前馈运动表征。讲话者直接读取运动指令的前馈控制能力，以及利用感觉反馈修正运动指令的反馈控制能力受到多种因素的影响。未来可以在以下几个方向进一步探索：

第一，在线言语运动控制的研究多数集中于相对稳态的单元音，如/a/或/u/ (Chen et al., 2012b; Scheerer et al., 2013a)。典型的音高听觉反馈扰动实验通常要求被试延长发音，夸大了元音的静态特征。然而，日常交流中的发音运动会导致声道形变，例如辅音和元音的过渡，因此语音的重要特征之一是会随时间不断动态变化。相比较而言，延长的静态发音几乎不会出现在自然连续语流中，因此生态效度较低。未来研究不应只关注简单的元音，也要关注自然语言，致力于更全面地理解言语运动控制的特性。

第二，从理论模型来看，虽然 DIVA 是言语运动控制领域现有最全面合理的模型之一，但也仍然存在一些缺点。首先，主要关注独立小单元的发音运动，如音节或经常使用的短语，忽略了言语运动系统是如何控制多音节单位中发音运动的过渡；再者，DIVA 模型也无法解释听觉反馈扰动实验中发现的泛化效应，即扰动特定语音的听觉反馈为何会导致相邻语音的言语运动调整；此外，心理语言学模型认为语言产生还包括概念确定、词条选择和单词形

式编码等步骤(Levelt et al., 1999)，Hickok (2012)指出 DIVA 模型更关注动力学、运动轨迹和反馈控制，忽略了与传统心理语言学视角的联系。最后，模型未关注言语产生中对韵律的控制，而音高、音强、音长和节奏传达了重要的语言学和情感信息(Tourville & Guenther, 2011)。目前 Guenther 教授团队开始发展 GODIVA(Gradient-order)模型，未来研究者也需要在更大的语言单位以及超音段水平上探索言语运动的控制机制。

第三，从影响因素来看，个体差异、训练经历和任务情境等，目前都被认为与前馈和反馈控制的整合加工密切相关，然而在该领域也有少数研究发现不统一的结果。此外，对于这些因素是如何影响或如何共同影响言语运动控制，远没有得到充分了解。未来研究应重点考察影响前馈和反馈控制整合加工的因素以及相应的理论依据。

第四，从语言类型来看，双语者第二语言运动控制的研究刚刚起步，与母语差异的探索亟待加强。Simmonds 等(2011)提出虽然母语和二语共享感知−运动脑网络，但理论上来说，母语中运动指令高度自动化，并高效整合来自前馈运动、听觉反馈和体觉反馈的信息，而二语中运动指令不熟悉，而且听觉反馈经常与内部表征不匹配，因而可能需要更多感觉反馈控制的参与。此外，晚期双语者普遍存在的口音问题实质上反映了二语中运动−感觉信息的整合很难达到母语水平。为此我们应开展实证研究揭示二语言语运动控制的一般性规律，为促进第二语言语音教学提供理论指导。

第五，探索言语运动控制的神经机制。对言语运动系统及认知机制的早期探索主要来自声学研究，近年来研究者尝试采用脑成像技术探讨听觉加工和运动修正发生的时间进程及其神经基础。但出声命名遇到最大的质疑是发声带来的运动相关伪迹，研究者认为运动伪迹可能会掩蔽由实验操纵带来的神经活动变化。鉴于此，部分研究者要求被试在语言产生过程中尽量减少发音器官的移动(Chen et al., 2012b)，或采用严格的数据剔除标准(Scheerer et al., 2013a, 2013b)，抑或采用不出声命名任务(Tian & Poeppel, 2015)。下一步研究应着重思考如何结合行为、ERP、MEG 和 fMRI 等技术以及多种实验范式，深入有效地考察前馈和反馈控制整合加工的认知神经机制。

参考文献

张清芳，杨玉芳．(2003)．言语产生中的词汇通达理论．*心理科学进展，11*(1), 6–11.

Alsius, A., Mitsuya, T., & Munhall, K. (2013). Does compensation in auditory feedback require attention? *Journal of the Acoustical Society of America, 133*(5), 3342.

Ballard, K. J., Halaki, M., Sowman, P. F., Kha, A., Daliri, A., Robin, D., ... Guenther, F. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Frontiers in Human Neuroscience, 12*, 510.

Ballard, K. J., Tourville, J., & Robin, D. A. (2014). Behavioural, computational, and neuroimaging studies of acquired apraxia of speech. *Frontiers in Human Neuroscience, 8*, 892.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America, 119*(4), 2363–2371.

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2015). Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control. *Frontiers in Neuroscience, 9*, 109.

Behroozmand, R., & Larson, C. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience, 12*(1), 54–63.

Behroozmand, R., Sangtian, S., Korzyukov, O., & Larson, C. R. (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research, 1636*, 1–12.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage, 32*(2), 821–841.

Cai, S. (2012). *Online control of articulation based on auditory feedback in normal speech and stuttering: Behavioral and modeling studies* (Unpublished doctorial dissertation). Massachusetts Institute of Technology, Cambridge.

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE, 7*(7), e41830.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production

of the formant trajectories in the Mandarin triphthong /iau/ and its patterns of generalization. *Journal of the Acoustical Society of America, 128*(4), 2033–2048.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience, 31*(45), 16483–16490.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414–421.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences, 110*(7), 2653–2658.

Chen, Z., Chen, X., Liu, P., Huang, D. & Liu, H. (2012a). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neuroscience, 13*(1), 1–10.

Chen, Z., Liu, P., Jones, J. A., Huang, D., & Liu, H. (2010). Sex-related differences in vocal responses to pitch feedback perturbations during sustained vocalization. *Journal of the Acoustical Society of America, 128*(6), EL355–EL360.

Chen, Z., Liu, P., Wang, E. Q., Larson, C. R., Huang, D., & Liu, H. (2012b). ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization. *Brain and Language, 121*(1), 25–34.

Chen, Z., Wong, F. C. K., Jones, J. A., Li, W., Liu, P., Chen, X., et al. (2015). Transfer effect of speech-sound learning on auditory-motor processing of perceived vocal pitch errors. *Scientific Reports, 5*, 13134.

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders, 35*(3), 246–279.

Cowie, R., Douglas-Cowie, E., & Kerr, A. G. (1982). A study of speech deterioration in post-lingually deafened adults. *The Journal of Laryngology & Otology, 96*(02), 101–112.

Cruikshank, L. C., Singhal, A., Hueppelsheuser, M., & Caplan, J. B. (2012). Theta oscillations reflect a putative neural mechanism for human sensorimotor integration. *Journal of Neurophysiology, 107*(1), 65–77.

Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language, 150,* 37–44.

Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language, 143,* 59–68.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2017). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2), e12521.

Delvaux, V. & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica, 64*(2-3), 145–173.

Franken, M. K., Acheson, D. J., Mcqueen, J. M., Peter, H., & Frank, E. (2018a). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review, 25*(4), 1458–1467.

Franken, M. K., Frank, E., Acheson, D. J., Mcqueen, J. M., Peter, H., & Jan-Mathijs, S. (2018b). Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. *NeuroImage, 179*, 326–336.

Franklin, D. W. & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron, 72*(3), 425–442.

Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro, E., Jr., Andrew, C. M., et al. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex, 16*(7), 969–977.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage, 52*(3), 862–874.

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage, 55*(3), 1324–1338.

Gould, J., Lane, H., Vick, J., Perkell, J. S., Matthies, M. L., & Zandipour, M. (2001). Changes in speech intelligibility of postlingually deaf adults after cochlear implantation. *Ear and Hearing, 22*(6), 453–460.

Grafton, S. T., Schmitt, P., Van Horn, J., & Diedrichsen, J. (2008). Neural substrates of visuomotor learning based on improved feedback and prediction. *NeuroImage, 39*(3), 1383–1395.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72*(1), 43–53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*(3), 594–621.

Guenther, F. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*(5), 350–365.

Guenther, F. H. (2016). *Neural control of speech.* Cambridge, MA: MIT Press.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280–301.

Guenther, F., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25*(5), 408–422.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology, 42*(2), 180–190.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport, 17*(13), 1375–1379.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135-145.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407–422.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology, 101*(5), 2725–2732.

Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology, 33*, 174–181.

Houde, J. F. & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5*, 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of cognitive neuroscience, 14*(8), 1125–1138.

Indefrey, P. (2011). The spatio and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101–144.

Ito, M. (2000). Mechanisms of motor learning in the cerebellum. *Brain Research, 886*(1-2), 237–245.

Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54,* 32–42.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research, 190*(3), 279–287.

Jones, J. A., Scheerer, N., & Tumber, A. (2013). The relationship between vocal pitch feedback error and event-related brain potentials. In *Proceedings of Meetings on Acoustics,* Vol.19, 060151.

Kakimoto, A., Ito, S., Okada, H., Nishizawa, S., Minoshima, S., & Ouchi, Y. (2016). Age-related sex-specific changes in brain metabolism and morphology. *Journal of Nuclear Medicine, 57*(2), 221–225.

Kalpouzos, G., & Nyberg, L. (2010). *Asymmetry of Memory in the Brain.* In Hugdahl K, Westerhausen R, eds. The two halves of the brain: Information processing in the cerebral hemispheres. MIT Press, Cambridge, MA, USA. 499–530.

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience,* 1–16.

Kearney, E., Nieto-Castoñón, A., Weerathunge, H.R., Falsini, R., Daliri, A., Abur, D., ...Guenther, F.H. (submitted). A simple 3-parameter model for examining adaptation in speech and voice production.

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience, 14*(1), 25–35.

Kort, N. S., Nagarajan, S. S., & Houde, J. F. (2014). A bilateral cortical network responds to pitch perturbations in speech feedback. *NeuroImage, 86*, 525–535.

Korzyukov, O., Karvelis, L., Behroozmand, R., & Larson, C. R. (2012a). ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. *International Journal of Psychophysiology, 83*(1), 71–78.

Korzyukov, O., Sattler, L., Behroozmand, R. & Larson, C. R. (2012b). Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback. *PLoS One, 7*(7), e41216.

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science, 25*(7), 1325–1336.

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience, 32*(27), 9351–9358.

Lane, H. & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America, 89*(2), 859–866.

Levelt, W. J. M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–75.

Li, J., Hu, H., Chen, N., Jones, J. A., Wu, D., Liu, P., & Liu, H. (2018). Aging and sex influence cortical auditory-motor integration for speech control. *Frontiers in Neuroscience, 12*, 749.

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology, 122*(5), 951–957.

Liu, H., Russo, N., & Larson, C. R. (2010a). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *Journal of the Acoustical Society of America, 127*(2), 1042–1046.

Liu, H., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., & Huang, D. (2010b). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during vocalization. *Journal of the Acoustical Society of America, 128*(6), 3739–3746.

Liu, X., & Tian, X. (2018). The functional relations among motor-based prediction, sensory goals and feedback in learning non-native speech sounds: Evidence from adult Mandarin Chinese speakers with an auditory feedback masking paradigm. *Scientific Reports, 8*(1), 11910.

Liu, Y., Hu, H., Jones, J., Guo, Z., Li, W., Chen, X., et al. (2015). Selective and divided attention modulates auditory-vocal integration in the processing of pitch feedback errors. *European Journal of Neuroscience, 42*(3), 1895–1904.

Maas, E., Mailend, M. L., & Guenther, F. H. (2015). Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language, and Hearing Research, 58*(2), 185–200.

Mitsuya, T., Munhall, K. G., & Purcell, D. W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *Journal of the Acoustical Society of America, 141*(4), 2758–2767.

Munhall, K. G., Macdonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America, 125*(1), 384–390.

New, A. B., Robin, D. A., Parkinson, A. L., Duffy, J. R., McNeil, M. R., Piguet, O., et al. (2015). Altered resting-state network connectivity in stroke patients with and without apraxia of speech. *NeuroImage: Clinical, 8*, 429–439.

Ning, L.-H., Loucks, T. M., & Shih, C. (2015). The effects of language learning and vocal training on sensorimotor control of lexical tone. *Journal of Phonetics, 51*, 50–69.

Ning, L.-H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and

sensorimotor contributions. *Speech Communication, 63-64*, 55–69.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33*(41), 16110–16116.

Oller, D. K. & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59*(2), 441–449.

O'Reilly, J. X., Mesulam, M. M., & Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience, 28*(9), 2252–2260.

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage, 61*(1), 314–322.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience, 37*(38), 9249–9258.

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *Journal of the Acoustical Society of America, 145*(3), 1456–1481.

Patel, R., Reilly, K. J., Archibald, E., Cai, S., & Guenther, F. H. (2015). Responses to intensity-shifted auditory feedback during running speech. *Journal of Speech Language and Hearing Research, 58*(6), 1687–1694.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics, 25*(5), 382–407.

Reilly, K. J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology, 118*(5), 2925–2934.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382.

Scheerer, N. E., Behich, J., Liu, H., & Jones, J. A. (2013a). ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience, 240*, 176–185.

Scheerer, N. E. & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience Letters, 529*(2), 128–132.

Scheerer, N. E., Liu, H., & Jones, J. A. (2013b). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience, 38*(8), 3189–3200.

Scheerer, N. E., & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *European Journal of Neuroscience,*

40(12), 3793–3806.

Scheerer, N. E., & Jones, J. A. (2018). Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors. *Neuropsychologia, 114*, 158–167.

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two tongues, one brain: Imaging bilingual speech production. *Frontiers in Psychology, 2*, 166.

Swink, S., & Stuart, A. (2012). The effect of gender on the N1-P2 auditory complex while listening and speaking with altered auditory feedback. *Brain and Language, 122*(1), 25–33.

Terband, H., Rodd, J., & Maas, E. (2015). Simulations of feedforward and feedback control in apraxia of speech (AOS): Effects of noise masking on vowel production in the DIVA model. In *The 18th International Congress of Phonetic Sciences.*

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology, 1*, 166.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in human neuroscience, 6*, 314.

Tian, X., & Poeppel, D. (2015). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive Neuroscience, 27*(2), 352–364.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *CORTEX, 77*(7), 1–12.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., et al. (2007). Neural correlates of auditory feedback control in human. *Neuroscience, 146*(2), 499–503.

Tumber, A. K., Scheerer, N. E., & Jones, J. A. (2014). Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *PLoS ONE, 9*(10), e109968.

Tye-Murray, N, & Spencer, L. (1995). Acquisition of speech by children who have prolonged cochlear implant experience. *Journal of Speech and Hearing Research, 38*(2), 327–337.

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience, 12*(12), 739–751.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience, 22*(8), 1770–1781.

The integration mechanisms of feedforward and feedback control in speech motor system

CAI Xiao; ZHANG Qingfang

(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract: The final stage in the process of spoken production is articulation, which involves the integration of feedforward and feedback control in speech motor system. Specifically, feedforward control (top-down mechanism) refers to speakers' ability to retrieve and execute the motor commands responsible for producing target speech sounds, while feedback control (bottom-up mechanism) refers to speakers' ability to adjust speech movements based on the sensory feedback generated by articulation. Sensory goals and sensory predictions are important hubs linking feedforward and feedback control systems. Based on the neural computational model DIVA (directions into velocities of articulators), the cognitive and neural mechanisms of the integration between feedforward and feedback control are illustrated in the stage of speech acquisition and speech production. On the basis of previous studies, how speakers utilize auditory feedback to control online speech and update feedforward motor representations, and the cognitive significance of the P1-N1-P2 components in the ERP studies are especially discussed. Furthermore, various factors that influence feedforward and feedback control are summarized, including individual variabilities, training experience and task demands. Additionally, some suggestions are proposed for future investigation.

Key words: speech motor system; feedforward control; feedback control; auditory feedback